Formale Grundlagen der Informatik I
6. Übungsblatt

TECHNISCHE UNIVERSITÄT DARMSTADT

Fachbereich Mathematik
Prof. Dr. Martin Ziegler
Alexander Kreuzer
Carsten Rösnick

SS 2011
18.05.11

Minitest Lösung

a) Betrachten Sie die Grammatiken $G_i = (\{a,b\}, S, X, Y\}, P_i, S)$ für $i \in \{1,2\}$ und mit

$$\begin{array}{llll} P_1: & S & \to & XY \\ & X & \to & a \mid aX \\ & Y & \to & b \mid bY \end{array} \qquad \begin{array}{llll} P_2: & S & \to & XY \\ & X & \to & a \mid aX \\ & Y & \to & b \mid Yb \\ & XY & \to & YX \end{array}$$

Welchen Typ haben die Grammatiken G_1 und G_2? G_1: ☐ 3 ☒ 2 ☒ 1 ☒ 0
(mehrere Antworten möglich) G_2: ☐ 3 ☐ 2 ☒ 1 ☒ 0

Begründung: G_1 ist Typ 2, weil jede Produktion der Form $X \to v$ für eine Variable X und ein v ist. Sie ist nicht Typ 3, weil $X_0 \to XY$ bei regulären Grammatiken nicht erlaubt ist.
G_2 ist Typ 1, weil jede Produktion nicht verkürzend ist. Sie ist nicht Typ 2, weil $XY \to YX$ mehr als eine Variable auf der linken Seite hat.

b) Welche der folgenden Sprachen beschreibt die Grammatik G_1 aus a) ?

☐ $\{a^n b^n : n \geq 1\}$ ☐ $a^* b^*$ ☒ aa^*bb^* ☐ $(a+b)^*$

Begründung: Aus X lässt sich genau a^{n+1} und $a^n X$ für beliebiges $n \in \mathbb{N}$ ableiten. Genauso lässt sich aus Y genau b^{m+1} und Yb^m für jedes $m \in \mathbb{N}$ ableiten. Damit kann man aus S genau $a^{n+1}b^{m+1}$, $a^{n+1}Yb^m$, a^nXb^{m+1}, a^nXYb^m für jedes $n,m \in \mathbb{N}$ ableiten. Also $L(G_1) = \{a^{n+1}b^{m+1} \colon n,m \in \mathbb{N}\} = L(aa^*bb^*)$.

c) Welche der folgenden Implikationen gilt im Allgemeinen?
Die Sprache L ist regulär.

☐ $\Leftarrow$ ☒ $\Rightarrow$ Es gilt die Aussage aus dem Pumping Lemma, d.h. es gibt ein n, so dass sich jedes Wort $x \in L$ mit $|x| \geq n$ in $x = u \cdot v \cdot w$ zerlegen läßt ...

☒ $\Leftarrow$ ☒ $\Rightarrow$ Die Relation $\sim_L$ hat endlichen Index.

Begründung: Nach Pumping Lemma, gilt natürlich die Aussage des Pumping Lemmas für reguläre Sprachen. Umgekehrt gilt aber nicht, dass eine Sprache regulär ist, wenn sie die Aussage erfüllt. Ein Gegenbeispiel haben wir in der Aufgabe G3 im 5. Übungsblatt konstruiert. Die andere Aussage ist genau die Charakterisierung aus dem Satz von Myhill-Nerode.

Gruppenübung

Aufgabe G1
Betrachten Sie die kontextfreie Grammatik $G = (\{a,b\}, \{X_0, X, Y\}, P, X_0)$ mit

$$\begin{array}{llll} P: & X_0 & \to & aXaY \\ & X & \to & aXa \mid Y \\ & Y & \to & bY \mid \varepsilon. \end{array}$$

(a) Bestimmen Sie die von G erzeugte Sprache.

(b) Bestimmen Sie eine zu G äquivalente Grammatik G' ohne ε-Produktionen.

Lösungsskizze:

(a)

$$L(G) = \{a^n b^m a^n b^k \,:\, n \geq 1, m \geq 0, k \geq 0\}.$$

(b) Die Menge der Variablen Z, die nach endlich vielen Ableitungsschritten das leere Wort produzieren, ist $V_\varepsilon = \{X, Y\}$ (vgl. auch Beweis von Lemma 2.4.3). Wir entfernen $Y \to \varepsilon$ und fügen dafür die folgenden Produktionen zu G hinzu:

$$\begin{array}{lcl} X_0 & \to & aXa \\ X & \to & aXa \,|\, \varepsilon \\ Y & \to & b. \end{array}$$

Weiterhin entfernen wir $X \to \varepsilon$ und fügen auch die nachfolgenden Produktionen zu G hinzu (dabei die eben neu zu G hinzugefügten Produktionen nicht vergessen!):

$$\begin{array}{lcl} X_0 & \to & aaY \\ X & \to & aa \\ X_0 & \to & aa. \end{array}$$

Nach dem Entfernen der ε-Produktionen erhalten wir insgesamt:

$$\begin{array}{lcl} X_0 & \to & aXaY \,|\, aXa \,|\, aaY \,|\, aa \\ X & \to & aXa \,|\, Y \,|\, aa \\ Y & \to & bY \,|\, b. \end{array}$$

Aufgabe G2

Betrachten Sie die kontextfreie Grammatik $G = (\{a, b\}, \{X_0, X, Y\}, P, X_0)$ mit

$$\begin{array}{llcl} P: & X_0 & \to & aXaY \,|\, aXa \,|\, aaY \,|\, aa \\ & X & \to & aXa \,|\, Y \,|\, aa \\ & Y & \to & bY \,|\, b. \end{array}$$

(a) Welche Sprache wird von G erzeugt?

(b) Konstruieren Sie eine Grammatik G' in Chomsky-Normalform mit $L(G) = L(G')$.

(c) Entscheiden Sie mit Hilfe des CYK-Algorithmus, ob das Wort $w = aabaa$ in der von G' erzeugten Sprache $L(G')$ liegt. Ist w in $L(G')$ enthalten, so geben Sie zudem eine Ableitung in G' an.

Lösungsskizze:

(a) Das ist immer noch

$$L(G) = \{a^n b^m a^n b^k \,:\, n \geq 1, m \geq 0, k \geq 0\}$$

(b) Erstens fügen wir Z_a und Z_b hinzu und

$$\begin{array}{lcl} X_0 & \to & Z_a X Z_a Y \,|\, Z_a X Z_a \,|\, Z_a Z_a Y \,|\, Z_a Z_a \\ X & \to & Z_a X Z_a \,|\, Y \,|\, Z_a Z_a \\ Y & \to & Z_b Y \,|\, Z_b \\ Z_a & \to & a \\ Z_b & \to & b. \end{array}$$

Dann eliminieren wir die $X \to Y$ Produktionen:

$$\begin{array}{lcl} X_0 & \to & Z_aXZ_aY \mid Z_aXZ_a \mid Z_aZ_aY \mid Z_aZ_a \\ X & \to & Z_aXZ_a \mid Z_bY \mid b \mid Z_aZ_a \\ Y & \to & Z_bY \mid b \\ Z_a & \to & a \\ Z_b & \to & b. \end{array}$$

Letztens eliminieren wir die $X \to X_1 \dots X_n$ Produktionen mit $n \geq 3$: wir fügen Z_1 hinzu und ersetzen $X_0 \to Z_aXZ_aY$ durch $X_0 \to Z_aXZ_1$ und $Z_1 \to Z_aY$, dann fügen wir Z_2 hinzu und ersetzen $X_0 \to Z_aXZ_1$ durch $X_0 \to Z_aZ_2$ und $Z_2 \to XZ_1$, fügen Z_3 hinzu und ersetzen $X_0 \to Z_aXZ_a$ durch $X_0 \to Z_aZ_3$ und $Z_3 \to XZ_a$, fügen Z_4 hinzu und ersetzen $X_0 \to Z_aZ_aY$ durch $X_0 \to Z_aZ_4$ und $Z_4 \to Z_aY$, und schließlich fügen Z_5 hinzu und ersetzen $X \to Z_aXZ_a$ durch $X \to Z_aZ_5$ und $Z_5 \to XZ_a$. Insgesamt:

$$\begin{array}{lcl} X_0 & \to & Z_aZ_2 \mid Z_aZ_3 \mid Z_aZ_4 \mid Z_aZ_a \\ X & \to & Z_aZ_5 \mid Z_bY \mid b \mid Z_aZ_a \\ Y & \to & Z_bY \mid b \\ Z_1 & \to & Z_aY \\ Z_2 & \to & XZ_1 \\ Z_3 & \to & XZ_a \\ Z_4 & \to & Z_aY \\ Z_5 & \to & XZ_a \\ Z_a & \to & a \\ Z_b & \to & b. \end{array}$$

(c) Der CYK-Algorithmus ergibt folgende Tabelle:

w $j\backslash i$	a 1	a 2	b 3	a 4	a 5
1	$\{Z_a\}$	—	—	—	—
2	$\{X_0, X\}$	$\{Z_a\}$	—	—	—
3	$\{X_0\}$	$\{Z_1, Z_4\}$	$\{X, Y, Z_b\}$	—	—
4	$\emptyset$	$\{X_0, X\}$	$\{Z_3, Z_5\}$	$\{Z_a\}$	—
5	$\{X_0, X\}$	$\{Z_3, Z_5\}$	$\emptyset$	$\{X_0, X\}$	$\{Z_a\}$

Damit folgt $w = aabaa \in L(G')$, denn das Startsymbol X_0 ist Element der Menge des Eintrags $(i, j) = (1, 5)$. Dabei leitet sich w bspw. wie folgt ab:

$$X_0 \to_{G'} Z_aZ_3 \to_{G'} aXZ_a \to_{G'} aZ_aZ_5a \to_{G'} aaXZ_aa \to_{G'} aabaa = w$$

Aufgabe G3

(a) Welche Sprache beschreibt die Grammatik $G = (\Sigma, V, P, S)$ mit $\Sigma := \{a, b, c, \$\}$, $V := \{S, A, B, X, Y\}$ und Produktionen

$$\begin{array}{ll} S \to AB & Xa \to aX \\ A \to aXA \mid \$ & X\$ \to \$Y \\ B \to bYB \mid \$ & Yb \to bY \\ & Y\$ \to \$c \end{array}$$

Beweisen Sie ihre Behauptung durch Induktion über die Menge der in G ableitbaren $(V \cup \Sigma)$-Wörter. Stellen Sie hierzu zuächst eine geeignete Induktionsbehauptung auf.

(b) Geben Sie eine kontextfreie Grammatik für die Sprache aus (a) an, d. h. eine Grammatik, bei welcher die linke Seite jeder Produktion aus einer einzigen Variablen besteht.

Lösungsskizze:

(a) Wir behaupten: $L(G) = \{a^k\$b^n\$c^{k+n} \mid k, n \in \mathbb{N}\}$.

($\supseteq$) Jedes Wort der Form $w := a^k\$b^n\$c^{k+n}$ kann in G abgeleitet werden:

$$\begin{aligned} S &\to_G AB \to_G aXAB \to_G aXaXAB \to_G \cdots \to_G (aX)^k AB \to_G (aX)^k\$B \\ &\to_G (aX)^k\$bYB \to_G \cdots \to_G (aX)^k\$(bY)^n B \to_G (aX^k)\$(bY)^n\$ \\ &\to_G \cdots \to_G a^k X^k\$(bY)^n\$ \to_G \cdots \to_G a^k\$Y^k(bY)^n\$ \\ &\to_G \cdots \to_G a^k\$b^n Y^{k+n}\$ \to_G \cdots \to_G a^k\$b^n\$c^{k+n} \end{aligned}$$

($\subseteq$) Wir beweisen die folgenden Aussagen durch Induktion über den Ableitungsprozeß.

i. Gilt $S \to_G^* w$, dann ist $w = S$ oder w ist von der Form $w = uxvyz$ für geeigenete Wörter

$$u \in \{a, X\}^*, \quad v \in \{b, Y\}^*, \quad z \in \{c\}^*, \quad x \in \{A, \$\} \quad \text{und} \quad y \in \{B, \$\}\,.$$

ii. Gilt $S \to_G^* w$, so ist $|w|_a + |w|_b = |w|_X + |w|_Y + |w|_c$.

Induktionsanfang: Für $w = S$ sind die Aussagen offensichtlich erfüllt.

Induktionsschritt: Angenommen die Ableitung ist $S \to_G^* w' \to_G w$. Falls $w' = S$, dann ist $w = AB$ und (1) und (2) erfüllt. Wir nehmen deshalb an, dass $w' \neq S$. Nach Induktionshypothese erfüllt w' die Bedingungen (2) und es gibt eine Zerlegung $w' = u'x'v'y'z'$, wie unter (1). Für jede mögliche Produktion, die von w' nach w führen könnte, müssen wir untersuchen, ob sie die Eigenschaften (1) und (2) erhält. Wir betrachten exemplarisch einige Fälle.

($A \to aXA$) In diesem Fall ist $x' = A$ und $w = u'aXAv'y'z'$. Desweiteren gilt

$$|w|_a + |w|_b = |w'|_a + |w'|_b + 1 = |w'|_X + |w'|_Y + |w'|_c + 1 = |w|_X + |w|_Y + |w|_c\,.$$

($Yb \to bY$) In diesem Fall ist $w = u'x'vy'z'$, wobei das Wort v aus v' entsteht, indem zwei Buchstaben vertauscht wurden. An der Anzahl der verschiedenen Buchstaben ändert sich nichts.

($Y\$ \to \$c$) Dann ist $w = u'x'v\$cz'$ mit $v' = vY$ und $y' = \$$. Es verringert sich die Anzahl der Y um 1. Dafür steigt die Zahl der c.

Die verbleibenden Fälle zeigt man analog.

(b)

$$\begin{aligned} S &\to aSc\,|\$X \\ X &\to bXc\,|\$ \end{aligned}$$

Hausübung

Aufgabe H1 (6 Punkte)

Gegeben sei eine Grammatik $G = (\{0,1\}, \{S, A, B, C\}, P, S)$ mit folgenden Produktionen:

$$\begin{aligned} P:\quad & S \to AB \\ & A \to 0 \\ & B \to 1AB \mid ABC \mid 1 \\ & C \to 1C \mid \varepsilon \end{aligned}$$

(a) Überführen Sie G in eine Grammatik G' in Chomsky-Normalform.

(b) Entscheiden Sie durch Ausführung des CYK-Algorithmus, ob das Wort $w = 001011$ in der von G' erzeugten Sprache $L(G')$ liegt. (Der CYK-Algorithmus sei dabei wie in der Übung in einer $|w| \times |w|$-Tabelle durchzuführen.) Ist w in $L(G')$ enthalten, so geben Sie zudem eine Ableitung in G' an.

(c) Konstruieren Sie einen Kellerautomaten, der $L(G)$ erkennt, und begründen Sie die Korrektheit Ihrer Konstruktion (beispielsweise durch Angabe des entsprechenden Satzes aus Abschnitt 4.1 des Skripts, dessen Beweis die konkrete Konstrukion eines PDA aus einer gegebenen Typ-2 Sprache aufzeigt).

Lösungsskizze:

(a) Schritt 1: Entfernen der ε-Produktionen. Füge dazu

$$B \to AB, \quad C \to 1$$

der Grammatik G hinzu. Wir erhalten:

$$\begin{aligned} P: \quad & S \to AB \\ & A \to 0 \\ & B \to 1AB \mid ABC \mid AB \mid 1 \\ & C \to 1C \mid 1 \end{aligned}$$

Schritt 2: Überführe die Grammatik in Chomsky-Normalform. Wir fügen dazu die Produktionen

$$D \to 1, \quad E \to AB, \quad F \to BC$$

hinzu und erhalten:

$$\begin{aligned} P: \quad & S \to AB \\ & A \to 0 \\ & B \to AB \mid AF \mid DE \mid 1 \\ & C \to DC \mid 1 \\ & D \to 1 \\ & E \to AB \\ & F \to BC \end{aligned}$$

(b) Die Tabelle des CYK-Algorithmus für das Wort $w = 001011$ ergibt sich wie folgt:

w $j\backslash i$	0 1	0 2	1 3	0 4	1 5	1 6
1	$\{A\}$	—	—	—	—	—
2	$\emptyset$	$\{A\}$	—	—	—	—
3	$\{B,E,S\}$	$\{B,E,S\}$	$\{B,C,D\}$	—	—	—
4	$\emptyset$	$\emptyset$	$\emptyset$	$\{A\}$	—	—
5	$\{B,E,S\}$	$\{B,E,S\}$	$\{B\}$	$\{B,E,S\}$	$\{B,C,D\}$	—
6	$\{B,E,F,S\}$	$\{B,F\}$	$\{F\}$	$\{B,F\}$	$\{C,F\}$	$\{B,C,D\}$

Damit folgt: $w \in L(G)$; eine Ableitung ist

$$S \to_{G'} AB \to_{G'} 0AF \to_{G'} 00BC \to_{G'} 00DE1 \to_{G'} 001AB1 \to_{G'} 001011.$$

(c) Sei $\mathcal{P} = (\Sigma, Q, q, \Delta, A, \Gamma, \#)$ der entsprechende Kellerautomat mit folgenden Transitionen (konstruiert anhand der Regeln aus dem Beweis von Satz 4.1.5):

$$
\begin{array}{l}
(q, \#,\ \varepsilon,\ S, \quad\ \ q) \\
(q, S,\ \varepsilon,\ AB, \quad q) \\
(q, A,\ \varepsilon,\ 0, \quad\ \ q) \\
(q, B,\ \varepsilon,\ 1AB,\ q) \\
(q, B,\ \varepsilon,\ ABC, q) \\
(q, B,\ \varepsilon,\ 1, \quad\ \ q) \\
(q, C,\ \varepsilon,\ 1C, \quad q) \\
(q, C,\ \varepsilon,\ \varepsilon, \quad\ \ q) \\
(q, 0,\ 0,\ \varepsilon, \quad\ \ q) \\
(q, 1,\ 1,\ \varepsilon, \quad\ \ q)
\end{array}
$$

Aufgabe H2 (Pumping-Lemma)
Zeigen Sie, dass die Sprache

$$L = \{a^n b^m a^n b^m \ :\ m, n \geq 0\}$$

nicht kontextfrei ist.

Lösungsskizze: Wir zeigen, dass das Pumping Lemma verletzt ist. Das heißt:

Für jedes $i \in \mathbb{N}$ gibt es ein $x \in L$ mit $|x| \geq i$, so dass für alle Zeichenreihen y, u, v, w, z, mit $x = y \cdot u \cdot v \cdot w \cdot z$, $|uw| > 0$ und $|uvw| \leq i$, es ein $j \in \mathbb{N}$ gibt mit

$$y \cdot u^j \cdot v \cdot w^j \cdot z \notin L.$$

Sei $i \in \mathbb{N}$ beliebig und betrachte

$$x = a^i b^i a^i b^i.$$

Offensichtlich ist $x \in L$ und $|x| \geq i$. Seien also y, u, v, w, z, mit $x = y \cdot u \cdot v \cdot w \cdot z$, $|uw| > 0$ und $|uvw| \leq i$. Wir wählen $j = 2$ und behaupten

$$x' = y \cdot u^2 \cdot v \cdot w^2 \cdot z \notin L.$$

Beweis: Nenne die Teilwörter der Form a^i und b^i in x *Blöcke*. Es gibt jetzt drei Möglichkeiten:

- Entweder u oder w überschreitet eine Blockgrenze (beides zusammen ist wegen $|uvw| \leq i$ unmöglich!). Dann enthält entweder u oder w sowohl a's als b's. Das bedeutet, dass x' nicht die Form $a^*b^*a^*b^*$ hat und deshalb nicht in L enthalten sein kann.
- u und w sind in demselben Block enthalten. Dann hat x' einen Block der länger ist als die andere drei und ist deshalb nicht in L enthalten.
- u und w sind in verschiedenen Blöcken. Diese Blöcke müssen benachbart sein, wegen $|uvw| \leq i$. Das heißt, dass entweder u nur aus a's besteht und w aus b's, oder umgekehrt. Dann haben die zwei a-Blöcke in x' verschiedene Länge (wenn uw a's enthält), oder die zwei b-Blöcke (wenn uw b's enthält; beides ist natürlich auch möglich!). Jedenfalls ist x' nicht in L enthalten.